DOCUMENT RESUME

ED 233 733 IR 050 380

AUTHOR Robbin, Alice
TITLE Strategies for Increasing the Use of Statistical Data. Occasional Paper Number 158.
INSTITUTION Illinois Univ., Urbana. Graduate School of Library and Information Science.
PUB DATE Apr 83
NOTE 33p.
AVAILABLE FROM Publications Office, 249 Armory Building, 505 East Armory Street, Champaign, IL 61820.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Business; *Databases; *Government (Administrative Body); Information Centers; Information Networks; Information Services; *Information Utilization; *Statistical Data; Technological Advancement
IDENTIFIERS Database Producers; Information Transfer; Machine Readable Data; *Numeric Data Bases; *Private Sector; Public Sector; University Affiliated Facilities

ABSTRACT

This paper provides a complete picture of the environment for numeric databases and a set of appropriate responses for modifying this environment. The emphasis is on the statistical database produced by the public and private sectors, with coverage of government and academic stakeholders in the statistical data transfer system. Part one describes the interdependent components of the environment--producers, distributors, users, resources, technology, and the infrastructure--and also the constraints that impede efficient and effective statistical utilization, including discontinuities in the infrastructure; fiscal stress; the competing values of privacy or access, openness or secrecy, and public and private sector responsibilities; levels of technology and expertise; and the legal-institutional framework. Part two proposes five macro-level strategies designed to modify and improve the current environment for statistical data utilization and knowledge production, including developing federal policy guidelines and standards for improving the quality of numeric data; improving the degree and quality of the interaction between government and the statistical data communities; maximizing capital investment in government-and privately-produced data resources; improving the quality of education for both information specialists and the end-users of statistical data; and altering the responsibilities of the public and private sectors for statistical data production, transfer, and utilization. The author's vita is attached. (Author/ESR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

University of Illinois
Graduate School of Library and Information Science

OCCASIONAL PAPERS

ISSN 0276 1769

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Number 158
April 1983

ED233733

Strategies for Increasing the Use of Statistical Data

by

Alice Robbin

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Earl Plested

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR050380

© 1983 The Board of Trustees of The University of Illinois

Contents

Abstract 3
Introduction 3
Strategy One: Gathering Environmental Information 4
Components of the Infrastructure 4
Constraints 12
Competing Values 16
Technology 18
Expertise 19
Legal-Institutional Framework 20
Strategy Two: Selecting Appropriate Responses 22
Improve the Quality of Numeric Evidence 22
Improve the Degree and Quality of Interaction Between Governmental and Statistical Data Communities 23
Maximize Capital Investment in Data Resources 24
Educate the Information Specialist and End-User 25
Modify Political Choices 26
Conclusion 27
Acknowledgments 27
References 27
Vita 30

ABSTRACT

Contrary to the recent library and information literature, sound decision-making for increasing the use of numeric data cannot be based on the narrow conceptual framework of applying business marketing techniques. Providing information services, no matter the medium, requires adopting *multiple* strategies because the information service is part of a large complex with unique information, political and economic properties. The first strategy necessitates gathering information about the environment and recognizing interdependencies of the component parts and system constraints. Selecting appropriate responses is the second strategy. The goal is a series of middle-range decisions capable of modifying the information transfer process and the infrastructure. Producing products and services is the third strategy. At this stage, knowledge of user characteristics and needs can be applied to developing appropriate tools and services. Business marketing techniques are useful for optimizing the exchange between the information facility and its clientele.

INTRODUCTION

This paper focuses on strategies one and two. It provides a complete picture of the environment for numeric databases and a set of appropriate responses for modifying this environment. The emphasis is on the statistical database produced by the public and private sectors, with extensive coverage of government and academic stakeholders in the statistical data transfer system. Part one describes the components of the environment: producers, distributors, users, resources, techology, and the infrastructure. These components are highly interdependent, particularly because statistical data are inextricably tied to computer technology and require the interactions of many different types of experts. The environment introduces a number of constraints which impede efficient and effective statistical utilization. These constraints include discontinuities in the infrastructure; fiscal stress; the competing values of privacy and access; public and private sector responsibilities; and openness and secrecy; technology; and the legal-institutional framework.

Part two proposes five "macro-level" strategies designed to modify and improve the current environment for statistical data utilization and knowledge production. These strategies are directed at influencing the system's capability and performance. These long-range strategies include developing federal policy guidelines and standards for improving the quality of numeric data; improving the degree and quality of the interaction between

government and the statistical data communities within and outside government; maximizing capital investment in government- and privately-produced data by utilizing existing library structures and information networks in the public and private sector; and altering the responsibilities of both sectors for statistical data production, transfer and utilization. Strategies for developing appropriate products and services and for administering an efficient and effective data service can then be designed.

STRATEGY ONE: GATHERING ENVIRONMENTAL INFORMATION

Pfeffer and Salancik note that "a good deal of organizational behavior, the actions taken by organizations, can be understood only by knowing something about the organization's environment and the problems it creates for obtaining resources.[1] They also suggest:

> What happens in an organization is not only a function of the organization, its structure, its leadership, its procedures, or its goals. What happens is also a consequence of the environment and the particular contingencies and the constraints deriving from that environment.[2]

This section discusses the characteristics of the environment of an organization providing numeric data services, including the resource of numeric data; the infrastructure of producers, distributors, users, and their interdependencies; and constraints embodied by the information system.

Components of the Infrastructure

Producers and Distributors

Government is the largest producer of statistical data, collecting data as an integral part of its management and accountability responsibilities and regulatory activities. Government has traditionally assumed the responsibility for the collection, distribution, and analysis of statistical data. The statistical policies and standards it establishes and methodologies it develops subtly and overtly influence public and private record keeping and statistical activities. The size and complexity of the U.S. federal governmental system create enormous problems in developing standards that ensure data quality and promote access and retrieval of statistical data.

The private for-profit sector has been an influential transfer agent for statistical data gathered by the federal government. It has reorganized numeric data in formats familiar to purveyors and users of bibliographic

databases. In some cases, it has compiled statistical data for target clients (e.g., both Citibank and Data Resources, Inc. market economic forecasting models).[3] Private nonprofit firms have also repackaged government-produced data, aiming for the most part at markets different from the private for-profit firms. For example, DUALabs has received government contracts to reorganize and develop software for the 1970 U.S. Census of Population and Housing, and a series of housing quality surveys, to name but two databases that this influential firm has marketed. Data repositories like the Inter-university Consortium for Political and Social Research (ICPSR) and other university-based centers have received grants to serve as depositories for governmental statistical and administrative records, which include health, income maintenance, vital statistics, employment, labor force statistics, and criminal victimization. These same nonprofit organizations have also acted as primary data producers, encoding statistical information contained in government publications (e.g., ICPSR has done this for all the decennial U.S. censuses of population and housing at the county level and also for roll call records of both houses of the U.S. Congress).

The private nonprofit centers have become depositories for private sector data-gathering activities as well. For example, the Roper Center at the University of Connecticut is the international repository for Roper and Gallup polls and many other commercial polling agencies. The State Data Center at the University of California at Berkeley archives the Field Corporation California Polls. The Louis Harris Data Center at the University of North Carolina, Chapel Hill, is the national repository for the Louis Harris polling agency. ICPSR is the depository for the CBS *New York Times* election surveys. Some of the acquisition and archiving of these commercial polls has been made possible through grants from the private and public sectors.

The private nonprofit centers also serve as depositories for statistical data produced by the scientific community. In fact, this was their primary mandate when most of them were first established. It should not be assumed, however, that the scientific community automatically deposits its data in these archives. Data, although primarily funded through governmental grants, are still assumed to be the property of the investigator who collected them.

Data distribution in the federal government takes place through the National Archives and Records Service, the National Technical Information Service, the Bureau of the Census, the National Center for Health Statistics, the National Center for Educational Statistics, and the Bureau of

Labor Statistics, to name the most prominent federal distributors. Some agency units distribute their data on an ad hoc basis when requests are made. But as a general rule, information on what these units produce and what contact people have within them is not easily located. To improve the situation, arrangements for distributing governmental data have been made by several agencies with the private nonprofit sector. Examples include the Bureau of Justice Statistics and the National Institute of Health (Institute on Aging) with ICPSR and the Bureau of the Census with DUALabs.

Finally, some of the statistical data labeled "governmental" are produced under contract by firms in the private for-profit or nonprofit sector. There are only weak links or no coordinating structures between them and officially designated governmental depositories. Some of these firms distribute data either on a profit or a cost-recovery basis. For example, the National Science Foundation (NSF) has had contracts with Moshman, Inc. and Westat to conduct national surveys of scientific and technical manpower which are then distributed by these two firms. The Data and Program Library Service, also a recipient of NSF funds, served as the archive for the data produced by the two firms in order to provide the scientific community with easier access to these files.[4]

The multiple centers for data production and distribution create severe problems for the development of standards for data quality, documentation, availability, costs, distribution, and preservation. These problems ultimately affect the use and usefulness of numeric data.

Users

David[5] and Robbin[6] have characterized users of statistical data as problem solvers—ones who identify a problem and employ complex strategies for solving it. Problem solvers develop testable hypotheses, locate appropriate data and statistical estimation techniques, retrieve and manipulate data, and derive facts useful for generating new data or producing reports necessary for decision-making. There are six different types of numeric data users:

1. the *fact-finder* frequently "encountered by the reference librarian...";
2. the *bottom-liner* and *trend-seeker* seeking "a set of facts or numbers in order to make generalizations about certain conditions or processes...";
3. the *negotiator-transformer* gathering and transforming numbers;
4. the *bottom-liner/trend-seeker* and *negotiator-transformer* with multiple responsibilities;

5. the *high priest* making "judgments about reliability, quality, and acceptable degree of error..."; and
6. the *scientist-sage* evaluating the "results of statistical analysis of numerical data and engaging in knowledge-producing information...."[7]

Many different specialists are engaged in the process of numeric data production, transfer, distribution, and use. Some develop the contents of the database and match information needs to the policy or research task at hand. Other people design the protocols or collect instruments, while even others are responsible for the actual collection or edit the information. Encoders translate the information into numeric form according to prescribed rules. Other specialists design the structure of the numeric information so that the computer can manipulate the information efficiently, while others select the statistical estimation techniques or algorithms to produce reports of the manipulation. Also included in this complex system are those who either interact with the computer or interpret the computer-generated reports. Descriptions of the numeric database require the expertise of indexers, abstractors, catalogers, and documentalists. For example, locating, acquiring and distributing numeric data require the skills of librarians. Data professionals must be able to identify the needs and strengths of their clientele, to educate users about the deficiencies of existing data and its documentation, and assist users in educating their colleagues. Obviously, this ability requires skills such as the knowledge of the nature of numeric data, how they are used, and evaluation techniques regarding quality issues and those associated with library management.

Resources

It is not an easy task to describe or classify numeric data. Their contents, quality, quantity, structures, and uses vary enormously. With few exceptions, their contents are defined by their intended uses, although the potential richness of numeric data is such that they can be used in ways never intended by their original producers.[8] Statistical information is used in all aspects of the decision-making process.

Numeric data are designed to represent some aspect or aspects of the empirical world. They may be the translation of wavelengths; chemical compounds; movement of water; pollution-to-clean air ratios; subjective responses to questions about politics, the economy or the quality of one's neighborhood; indicators of economic growth or decline; rates of inflation; consumer expenditures; population characteristics; the voting behavior of citizens or legislatures; inteligence quotient test scores of high school and elementary students; sucess or failure in an education system; etc.

Governments, for record keeping and accounting purposes, produce vital statistics: information on the eligibility status of income recipients, retirement benefits provided through Social Security, corporate statistics, sources of income, revenues and expenditures, and tax records. For example, declining birth rates, longer life expectancy and rapid migration have led to population redistribution. As a result, private individuals, firms and government agencies require timely, reliable population data.[9] Health delivery agencies assess community needs, firms project local product demands; and demographers study migration patterns.[10] Chemists require identification of compounds. Electric and gas companies forecast electricity and gas consumption.[11] Urban policy specialists examine the relation between the labor force migration and the marketplace.[12] Measurement of export prices is essential for understanding many important international trade issues. The federal government and businesses, therefore, need indexes such as producer, unit value and export prices.[13] Since higher education must be able to predict the likelihood of any significant decline in enrollment for colleges and universities, they must project college and university enrollments through the year 2000.[14] Local crime statistics have been recognized as deficient; thus, the U.S. Bureau of Justice's statistics have, for almost a decade, supported the collection of crime surveys to study the dimension, nature and impact of crime in the United States. These surveys provide a more accurate means of measuring the incidence and severity of criminal victimization.

The quantity of computerized numeric data produced by the public and private sectors worldwide is unknown. Every nation in the world produces machine-readable data. Examples of the types of machine-readable data (MRD) include vital records, individual and corporate income and earnings for every taxpayer, wage earner or corporation. The entire 1960, 1970 and 1980 U.S. censuses are in machine-readable form. Public opinion surveys, voting records of legislators, medical records, hospital records, sales, financial records of private and public banks and corporations are in machine-readable form. Information on land use has been computerized.[16] Social services' agencies have implemented database management systems to provide client and agency accountability and eligibility status information. Original manuscript records have been encoded by historians engaged in social history projects.

New and different data products can be easily derived from records collected for one purpose because of the relationship to the computer. For example, statistical sampling techniques can be used to create byproducts of administrative records, which can be linked to different data sources to create records different from the original, primary data. One example is the

Social Security Administration project to link the current population surveys with the Internal Revenue Service and earnings records.[17] Sampling of the Wisconsin database of more than 175,000 cases of income maintenance recipients produced a public use sample of 3500 cases, to be used by government and social researchers for program planning, evaluation and research in ways that the original administrative records were not designed. (The database management system was originally designed to determine eligibility and grant allocations for the income maintenance programs of Aid to Families with Dependent Children (AFDC), food stamps and medical assistance. By using the sample population, researchers can link tax and other administrative records, and study patterns of family finances and the effects of "Reagonomics" on the lives of poor people.)

Numeric data may be collected only once or may be updated irregularly or regularly. There are longitudinal panel studies of the labor force such as the *Panel Study of Income Dynamics*, the *National Longitudinal Surveys of the Labor Market*, the *New Jersey Income Maintenance Experiment* (also conducted in Denver and Seattle); studies of high school graduates such as the *National Longitudinal Survey of the High School Class of 1972* and *High School and Beyond* (the high school class of 1980); and longitudinal studies of the voting electorate such as the *American National Election Studies*.

Because numeric data reflect the empirical world, their structures are varied and often complex. Consider for a moment gene patterning, consumer expenditure patterns, or census tracts which all have many diverse households. The relationships embodied by entities are complex and data structures must accurately reflect these relationships (see fig. 1).

Figure 1 indicates the relationship of each of the individuals in a household to three income maintenance programs—AFDC, food stamps and medical assistance. Located in county number one are a number of households, of which only one's relationship is completely described (household number two). Household number two contains three people. Person number one is eligible for and receives money from the three income maintenance programs; person number two receives aid from AFDC and food stamps; person number three receives aid only from the medical assistance program (he or she is most certainly elderly or handicapped). To be eligible for the income maintenance programs, recipients are required to provide information on their assets and income. Vehicles are considered an asset. Vehicles number one and number two are owned by person number one; vehicle number three is owned by person number three.

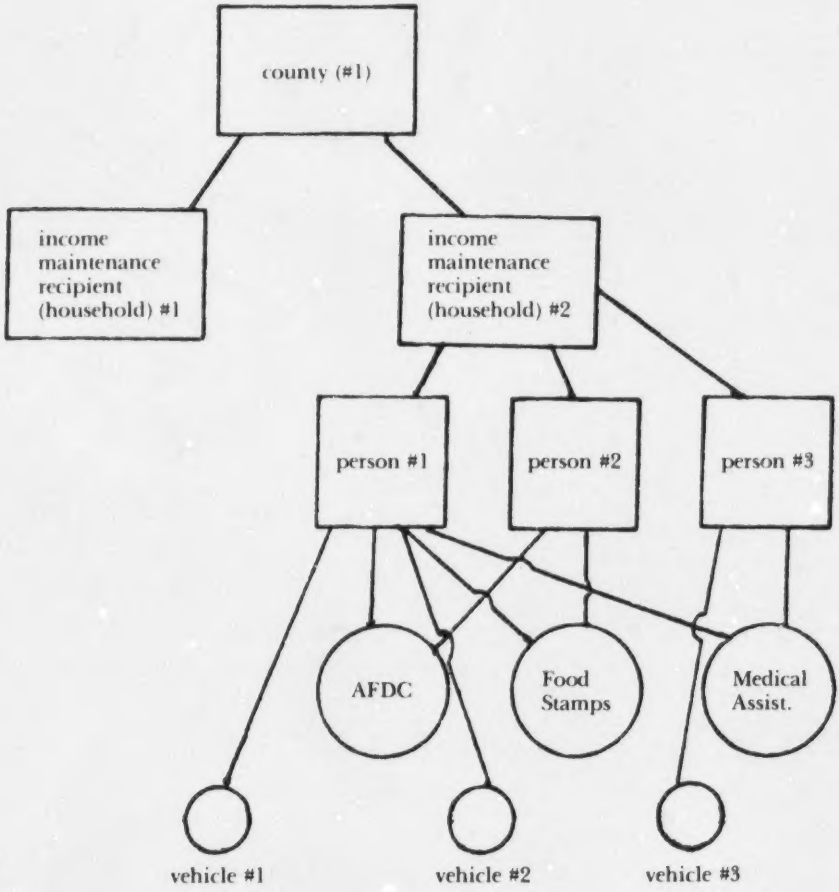

Fig. 1. An Example of a Complexly Structured Data Set

Computerized numeric data depend on descriptive materials called documentation for their interpretation. This documentation provides information on the reasons for their collection, the processing they underwent and the location and meaning of the file's contents.[18] Documentation also provides insights into the quality of the database.

Quality is a major theoretical and methodological issue. Lack of consistency in concepts means that the data may not be able to be used for comparative purposes, or that the conclusions drawn from the data might be inconclusive. Limitations in the data may seriously affect the ability of government and the private sector to plan programs, evaluate those programs, or forecast new developments. For example, until standardization of industrial and occupational classification occurred, regional policy makers in the European Economic Community were unable to examine the employment situation in the Common Market.[19] The Bureau of the Census, Social Security Administration and Internal Revenue Service are troubled by difficulties encountered with missing information, for it weakens the strength of the conclusions that can be drawn from their data.[20] Analysis of social service programs has suffered because of poor client records. Yet, accountability is a necessity, and good data are necessary for determining whether these programs have met their objectives.[21] For example, Dun and Bradstreet's *Dun's Market Indicator* file was believed to contain a reasonably complete listing of commercial and industrial establishments until a critical examination identified inaccuracies that would affect the quality of sample designs.[22] The Michigan Department of Treasury no longer publishes data on sales tax collections by county because their investigations showed these data were poor.[23] The poor quality of the sales tax data could jeopardize state budgetary forecasts. The large-scale migration that took place in the United States in the latter half of the 1970s meant that reliance on the 1970 U.S. Census of Population and Housing created significant forecasting problems for communities and businesses. As a result, many organizations seriously underestimated demands for services and products. Data quality will continue to be a major issue as demands for improved accuracy in forecasting and more responsiveness to social and economic problems are articulated by the public and private sectors.

Technology

Numeric data are inextricably bound to computer technology. This technology creates the possibility of efficient storage, retrieval, access, description, manipulation, management, organization, and reorganization.

The technology has evolved far more quickly than had been thought possible even less than a decade ago. Today, massive quantities of information can be stored and accessed at a fraction of the cost of five years ago. Indeed it is no longer the cost of the hardware that is the central problem faced by numeric data users, but rather the human resources required to develop and maintain the data, software and hardware which has become the central concern of the 1980s.

To this dynamic technology must be added communications technology. Computer and communications technologies have joined together to alter the speed by which information is transferred, the person who has access to the information, and the cost of both information and access. Just as computer storage technology has evolved to store and distribute information according to applications and user needs—reflecting the ability to modify hardware to special uses—the communications technology has made it possible to decentralize the storage of information. Now it is possible for many institutions and individuals either to rely on a central facility for storing and accessing their information while located far from this central facility, or to create specialized "nodes" in a decentralized information system, switching users and information among the nodes according to system capacity and user needs.

Constraints

Information systems do not operate in isolation, but within political, social and economic contexts. Information organizations, which are primarily service facilities, do not, as a rule, control the allocation, access and use of their resources—a control which is critical for survival of the operation. Networks and tools depend on the relationship of members of the information system to policy makers who provide political, economic and institutional support for the development and continuing support of an information infrastructure, and appropriate human capital and technology. Numeric data production, availability and utilization are dependent on the quality of political, legal, institutional, economic, and technological contexts. How these contexts have constraints is the subject of this section.

Infrastructure

Most producers of machine-readable data files have neither the resources nor expertise to preserve, maintain and distribute their data. In general, procedures for easy access to, and inexpensive use of, their data have not been established. Few data producers can supply adequate user support services related to the academic functions of teaching, research, coordination or multiple research projects, computation, and reference services.

Private for-profit sector activities, for the most part, have been designed to capitalize on repackaging government-produced data, targeted at an audience whose principal uses of the data are for increasing markets and forecasting economic trends. Their products are expensive and out of reach to all but the relatively wealthy corporate client. Secondary information services, patterned after bibliographic utilities, are few and costly, and

training in their use is limited to learning access and retrieval protocols. The contents—the quality and the unique nature of numeric data or statistical estimation techniques—and computer algorithms used to generate the reports are not emphasized. Government agencies, the largest producers of numeric data, with few exceptions, have major problems of production, quality control, coordination, dissemination, and description of their administrative and statistical products because these products have been designed principally for in-house tasks of record keeping, program planning and evaluation, with little or no intended use by consumers outside the government.[24] For this reason documentation is a rare occurrence. The formal links between government and the private sector have been limited, and government has not exercised much control over standards development for improving the quality of numeric data produced by either its own agencies or the private sector.[25]

In one sense, machine-readable data files have qualified as fugitive materials, unknown except to the cognoscenti. Serendipity—or accidental discovery of information—is a phenomenon known to all consumers and purveyors of information. But, for the numeric data user and data services professional, it is a common and daily occurrence. Informal channels of communication comprise the major means of access to information. In her study of the University of Wisconsin at Madison's Data and Program Library Service, Kathleen Heim showed that more than 60% of the university students learned of the existence of numerical data from a faculty member, colleague or friend. Less than 30% learned about data files from reports, journal articles, monographs, or consultation of the data library to search its directory and abstracts.[26] Faculty relied even less on the data library staff, and relied almost totally on a well-developed communication network of colleagues working on similar problems.

The data production, transfer and use process require integration with computational and statistical facilities. For the most part (with some important exceptions)[27] there are no formal communications channels for acquiring software and information on statistical computing. As a result, the consequences that follow include: (1) unavailability of reference consulting services for software acquisition and use, (2) duplication of effort in software creation, (3) lack of interest in program consulting staffs in the nature of the numeric data (and thus the relationship between the contents of numerical data and the application of correct statistical estimation techniques), and (4) an inability to exploit data resources because they require special software and computational facilities. Robbin surveyed users of scientific and technical manpower data resources. She found that a major problem associated with use of those data was the lack of appro-

priate computational facilities and staffs to assist in processing these large files.[28]

A key structural factor explains the lack of integration of numeric data in the library setting. Centers for statistical data—called data libraries—grew up outside the traditional library, as an adjunct of social science departments. Specialists in providing these services were not trained in information management, but in a discipline related to the information they disseminated. Their skills and training reflected their clientele's orientations. With few exceptions,[29] their clientele resisted integrating these special centers into the university library framework.[30] As a consequence, there have been (1) no transmission of professional library management skills into the data library, thus, organization of access to description of numeric data suffers; (2) no integration of numeric data into the general library, with one result being that little familiarity fosters disinterest on the part of librarians in this information resource; and (3) no description of numeric data resources with the result being that the availability of numeric data to the user community is known only to a very small number of potential consumers of numeric data services. Without centralizing description, machine-readable data files cannot enter the formal information system. Without catalog entries, users may not know how to cite numeric databases properly in their publications. By not entering the formal information system, abstracting and indexing of numeric data does not occur.

Fiscal Stress

Governments everywhere are under pressure to scale down or eliminate a number of censuses, surveys and administrative record keeping systems in order to create savings in programs and reduce reporting burdens. This reduction in funding levels has important political, economic, and perhaps, intellectual consequences.[31]

The Bureau of the Census is curtailing many economic and agricultural censuses, industrial reports, and educational and fertility surveys. The Social Security Administration is considering ending the coding of W-2 tax withholding forms for occupation and industry of workers, thus destroying the framework for much economic data. The Federal Trade Commission has threatened to end its quarterly financial reports, which are the source of corporate profits for the national income accounts. A study funded by the National Science Foundation which followed the economic and social lives of 5000 families for 14 years (*Panel Study of Income Dynamics*) has been terminated (although some funding was later provided by private foundations). Funds to support private sector archiv-

ing and distributing of federal statistical data (through data repositories) have been eliminated, reduced or held up.[32]

There are serious implications for information transfer, more particularly the end product that information services provide—public and private decision-making. For example, statistical data comprise the basis for determining eligibility for, and allocation of, funds for domestic assistance programs, such as health and income maintenance.[33] At a time when the Reagan administration is moving to devolve responsibilities for some social programs to the states and to introduce new economic and social policies, it is also losing its capability to monitor the effects of the Reagan economic plan on individual families. According to Joseph Duncan, former director of the Office of Federal Statistical Policy and Standards and now chief economist of Dun and Bradstreet: "In the long run...[there will be] serious trouble with...information on basic economic activity."[34]

For businesses, it "will mean less accurate data about important new industries and productivity trends...."[35] John Casson, a vice-president and chief economist of American Express, believes that the reductions in government programs will increase the likelihood of business and government decisions being based on erroneous information.[36]

The effects of reduction in statistical information production go beyond the loss of any particular series. In an interdependent system, funding reductions have spillover and feedback effects that are not immediately felt, but have long-term consequences for the economic, information and knowledge systems as a whole. Reductions will have subtle effects, such as smaller, more out-of-date samples, less quality control, less frequency of publication, decreased availability of government documents, and less research into improving the methodology of data collection and error reduction.[37] "Reductions in force" reduce the number of qualified personnel available to declassify, process and release information. The scientific community has become dependent on federal funding for large-scale investigations involving data collection and analysis. These investigations were designed to contribute to increasing the storehouse of knowledge about social, political and economic issues; evaluating social programs; and recommending solutions. Reductions in funding will most certainly affect data use in terms of producing new data. It may also negatively affect the use of existing data because the funds for data acquisition and maintenance, capital equipment purchase and teaching are no longer available. The private for-profit sector, which relies heavily on federal statistical data production, may have fewer products to market. With consumers affected by unemployment and inflation, there may be reluctance to invest discretionary dollars in information products.[38]

Competing Values

Public or Private Sector Responsibilities

Reductions in statistical information have made more salient the old debate over who shall be responsible for, and bear the costs of, information production and utilization—the public or private sector—and what should be the mix of responsibilities. Traditionally, the cost of producing public information was shared by all, without regard to who or how many used the information. Statistical data services, particularly in the private nonprofit and public sectors, absorbed most of the cost of data production. Minimal costs were passed on to the user, usually in the form of the direct costs of duplication of the data and documentation and of computer time. Neither staff time for providing the service nor overhead was built into the cost algorithm. But by the mid-1970s, most private nonprofit data services assessed the situation and began factoring in staff time and overhead. But, by and large, statistical databases have been a low cost burden for the user, consistent with the philosophy that statistical data should be available at as low a cost as possible to encourage their use. This decision was in sharp contrast to that made by private for-profit firms in the information industry which understood, quite properly, that information was an important and profitable commodity.[39]

Reductions in funding for statistics production, the Reagan administration's commitment to devolving information responsibilities to the private sector, and the restructuring of responsibilities for statistical policy (e.g., elimination of the Office of Federal Statistical Policy and Standards), raise a number of important questions regarding the mix of public and private sectors' responsibilities. There is still a recognition of and support for government to carry out certain functions that relate to the public interest. The issues are: (1) what is the public interest, (2) how much reliance should there be on the public sector for meeting statistical needs of the public, and (3) how much competition, if any, should there be betweeen public and private sectors?[40]

Privacy or Access

Statistical data comprise records of human processes, such as birth trauma, communicable diseases, and physical and psychological disabilities. They record matters which are private, yet which become "public" once government is charged with the responsibility for attending to the well-being and quality of life of its citizenry. Information of a private nature describes in detail individual attributes, and carries with it the capacity for intrusiveness and misuse. At the same time, this information is useful as a vehicle for examining the quality of government services; for research by the scholarly

community; and for policy and program planning, development and evaluation.

For more than a decade, lawmakers and citizens have grappled with the issues of confidentiality and access. Concern that the release of statistical data might compromise personal privacy has resulted in legislation restricting access to these data. At the same time there has been public policy activity designed to create greater openness in government, resulting in the enactment of open records legislation (e.g., U.S. Freedom of Information Act of 1966, and Amendments of 1974).

The difficulty in balancing these competing values has affected access to personally identifiable information by the social research community.[41] Increasing the use of statistical data is therefore intimately connected to availability, which is the subject of public policy debate. The real question is not *how* we will make this information available, but *whether* we will. In other words, it is a decision about social values, and a question of political choice. This is not to suggest that the dilemma cannot be resolved. Administrative, procedural and statistical strategies can be applied which affirm the privacy value as the preeminent one, while at the same time permitting access to records for social research and statistical purposes.[42]

Openness or Secrecy

In a democratic society, governments are accountable to their citizens. At some point, the information they gather must be available for public scrutiny. The debate revolves around what information should be public and when. Public policy discussions in 1974 resulted in the revised Freedom of Information Act. The act was designed to place public policy firmly on the side of more openness in the executive branch of government. One consequence of the legislation was administrative compliance and procedures for increasing the availability and use of government records.

During 1981-82, the Reagan administration attempted to "correct what it regard[ed] as excesses of open government spawned by the law."[43] Intensive lobbying was able to defeat several proposed amendments to the law. However, an Executive Order on Classification promulgated in spring 1982, gave officials broader authority to withhold information on the grounds of national security. It would not be unheard of if, for example, this Executive Order were interpreted to extend to public opinion surveys conducted by the Agency for International Development or the Voice of America.

Technology

Some of the characteristics that increase the utility of the numeric databases also create identification and copyright problems. Since numeric data are easily reformatted, updated, copied, and diffused through numerous agencies, the questions of which file or version of a file is the original and who has created that file become obscured.

The technology has changed dramatically. Computers produced less than fifteen years ago are no longer available. A numeric database dependent on an earlier computer model may not be usable on contemporary machines or even transferable from one computer site to another (without great expense). Data dependent on a particular and unique software may not be usable without that software; independence of the data from the software may be possible only at great cost. The growing complexity and scope of international and nationally-distributed statistical software packages is also a problem. Although these software packages are well-known and heavily used, many people may be unaware of the consequences of making mistakes in using them. Further, the algorithms for performing the same statistical estimation technique may be different. Users are often unaware of the strengths and weaknesses of the particular software packages. Specially created software with limited use may not be sufficiently tested on enough databases and by enough users to detect errors or potential problems.

Another factor is that the computer technology is evolving so rapidly that there are new products on the market every six months. Ironically, the virtually unlimited range of products has created a dilemma for the consumer because it becomes difficult to make a choice. More and more equipment is being designed for special-purpose applications targeted at a particular consumer group. But libraries, which need multiple types of application software for bibliographic control, statistics, report generating, account information, and record management, find it difficult to identify one computer that will meet all their needs and, perhaps more importantly, software that can be purchased "off the shelf." Libraries, at the current time, must still dedicate a significant percentage of their budget to programmers in order to create the special applications of software they need.

Current storage media and most of the mass storage devices that will soon be on the market do not meet low-cost, long-term archival storage and retrieval requirements. (The situation may soon change, however; but the cost of the long-term storage devices may place them out of reach of most data libraries.) Most large and small data files rely on magnetic tape as

backup (for protection, a copy is made of the data), and when not in use[44]—because the cost of the storage in the computer is too great and the magnetic tape is more "stable" than the computer where "crashes" could destroy mass storage files. (Storage on hard disk is not an option affordable by many data libraries, although hard disk offers a rather high quality storage medium.) Nevertheless, magnetic tape is not an enduring medium. Investigations here at the Data and Program Library Service indicated that all tapes produced before 1972 had to be discarded because of errors in reliability of the encoded information, and that there is a 20% loss of magnetic tapes from five to ten years old due to either poor quality control of the magnetic medium during production or poor environmental conditions.[45] In contrast to other media such as paper and fiche—whose life appears far longer than magnetic tape—a regular program of maintenance and preservation must be instituted. This contributes to the hidden costs of maintaining a library of numerical data.

The communications technology offers the potential of "distributed" responsibilities for preserving and providing access to special collections of data. The rate structure is changing, creating a more affordable way of reducing the need to maintain a copy of the data file in a library's collection. Increased transmission speeds are making it possible to transmit large quantities of statistical information at costs far lower than only a few years ago. However, the rate structure and transmission speeds are still not within the range of computing costs that most social scientists can afford. Further, it may still be necessary to maintain a copy of the data file in one's collection as further protection (that backup copy discussed earlier), as a means of reducing user data costs, and, at times, of creating efficiencies in retrieving the data. (It is currently easier to gain access to a local computer than one located at great distance, if only because use of a foreign computer requires more bookkeeping and knowledge of that computer's operating system. This situation will change, and, indeed, is already changing, as improvements are made in electronic accounting systems and in operating system software that is more "user friendly.")

Expertise

Users vary in the extent to which they understand theory, measurement, statistics, and computer science. There are also two other elements in the system that affect the ability to increase the use of numeric data—the data producers and the information specialist.

Extensive discussion of the need for government to review the quality of its statistical personnel has already been well documented.[46] Only a short

summary is necessary here. Professional training in data handling is received (or not received, as the case may be) on the job with little influence by nongovernment sources of expertise. The social science community, which has discovered many useful tools for improving data quality,[47] has limited opportunities for interacting with the government data producer and statistician. This interaction is neither encouraged by government and the university organization nor by attitudes of the government administrator or academician. Civil servants' opportunities for career development and participation in conferences are limited. In essence, holders of outdated information in a dynamic environment introduce inefficiencies in the statistical system. Lack of appropriate and current expertise has a direct influence on all aspects of the statistical data transfer process.

Library schools, one critical vehicle for training specialists in the intellectual and technical control and transfer of information, have been slow to reorient their structures of learning and communication to introduce concepts such as the interdisciplinary nature of statistical information, analytic techniques for evaluating information, and the integration of technology and information that could result in new information and new knowledge. Formal training of future library professionals in numeric data services is not institutionalized. Formal training in information management by numeric data center staffs is minimal or nonexistent. Institutionalized data professional structures are few and are not integrated into the larger network of information professional organizations.

Legal-Institutional Framework

Communications technology has made it possible to transmit and receive electronic information within seconds and at a relatively low cost. The technology is advancing so rapidly that increases in the capacity of the networks are occurring, while the costs of transmission are quickly decreasing. This is an important development for the statistical scientific community because the technology provides the capability of communicating beyond institutional and national boundaries. Yet, neither our legal system nor institutional framework has been sufficiently modified and reorganized to reflect this new reality (this is one of the discontinuities which I mentioned earlier). There are real costs and inefficiencies to the information system as a whole, and to the individual consumer, when laws and institutions are not flexible enough to incorporate technological change which modifies existing economic and social arrangements. Institutional rules and regulations, perhaps even laws, often prohibit the expenditure of money outside the user's institution, even when services can be acquired more inexpensively or efficiently. For example, most universi-

ties resist efforts to obtain services from an outside agent when similar services can be obtained from an internal unit or when a bid has been let to provide those services. So, if the local computing center provides data entry or data processing services, permission to obtain those services from outside the institution may be denied even when thay can be bought more cheaply.

A second way that the legal-institutional framework constrains organizations is in the purchase of new technologies. I have already described how the technology is changing so rapidly that every six months new equipment is available. Yet, the bidding process is proscriptive (usually specifying a particular model) and inefficient (responding to an already outdated market situation). An additional problem is that authority for designing specifications for bids and purchase is typically lodged in an administrative unit far from the actual consumer of the technology. Thus, the consumer has little effect over decisions made at a higher level, although he will ultimately be the one using the technology.

The legal-institutional framework constrains the information organization in subtle ways as well. The most important way, particularly for numeric data services that operate outside the traditional (university) library structure, is that these services are considered "support" or "auxiliary" facilities. This is in contrast to the central library whose data service is considered a "line item" in a budget (and may, as at the University of Wisconsin-Madison, have the second largest budget and the largest user community of all the units in the university). The data service library is dependent for its budget on either a single department or on the leadership of a few individuals (e.g., in a university environment the chief fiscal officer—the Dean of a College). Further, at some institutions, funding for special libraries is considered a separate item for negotiation (e.g., not part of the central library budget).[48]

The institutional framework may inhibit informal exchanges among library reference staffs, especially if formal communications channels have not been established. This is a particular problem for numeric data services because they have not, as a rule, been staffed by librarians. Although data services staffs might profit from interactions with other library staffs, this rarely occurs. Indeed, there is, in general, resistance on the part of data services professionals to accept the fact that their functions resemble very closely those of a librarian. Social networking, which is an important dimension to information provision, usually will not occur unless the institution provides mechanisms to facilitate this networking.

STRATEGY TWO: SELECTING APPROPRIATE RESPONSES

At the second stage in decision-making, holistic and process-oriented strategies must be developed if there are to be multiple outcomes.[49] A holistic approach focuses on efforts to improve the system's capability and performance, rather than on particular institutions or technical problems. Part two discusses this intermediate stage for developing strategies to increase the use of numeric data. These "middle range" strategies follow directly from the description of the environment presented here.

Improve the Quality of Numeric Evidence

Improving the quality of statistical evidence requires authoritative action at the highest level of policy making—the federal government. The decentralized federal statistical system requires a unifying institutional capacity to recommend and enforce standards for a whole range of actions. These include standards for (1) expertise in data production, transfer and use; (2) design, collection, production, and processing of data;[50] (3) description to increase access to, and use of, numeric data;[51] (4) uniformity in pricing structures to ensure equity and eliminate capriciousness and uncertainty by administrative staffs; and (5) institutionalized structures to provide and advise on data.

Implied in the general strategy is a commitment to public access to federal data. Problems with data collecting instruments, consistent time series, sample design, a lack of complete documentation, standardized coding schemes, and restrictions on confidential data can only be responded to if there is a commitment to making these data fully available in the public domain. Difficulties in locating and obtaining data due to high costs and confidentiality restrictions impede a full and open debate on the utility of statistical series and limit the ability to offer competing views on solutions to problems in contents, methodology and policy which the data were designed to answer.

A central statistical office of policy and standards is critically important for policy development and setting of agendas. Without it, elements in the decentralized federal statistical system are without direction. Many of the problems described in this section can be traced to the fact that the coordinating and overseeing functions of the multifaceted parts of the data system must be strengthened. As of this writing, there is no longer a coordinating body with a broad vision of the needs and capacities of the federal statistical system. Abolishing the Office of Federal Statistical Policy and Standards does not bode well for the future of the larger information

system, nor for political and economic strategies which depend on statistical data as a basis for decision-making.

Improve the Degree and Quality of Interaction Between Governmental and Statistical Data Communities

The information system in which statistical data production and analysis takes place is highly complex and dependent on new technologies. It requires expertise from many disciplines and specializations. The social scientist and policy maker have many common interests and a great deal to gain by cooperating to improve the quality of data and access to computerized statistical records. Government agencies must make greater use of available expertise in data collection and analysis activities, starting at the design stage and continuing through to an evaluation of how the results are used. More work needs to be carried out in the methodological area. Social scientists can contribute by conducting research into measurement of errors for improving collection methods and the presentation of information about methodological structure. Social scientists can develop a "consistent conceptual framework or model based on behavioral relationships in various disciplines...."[52] They can also develop standard concepts, definitions, classifications, survey frames, and procedures. Finally, social scientists can promote further utilization of data resources on behalf of the government and the private sector.

Through their activities, social scientists can promote record linkage at the microlevel and demonstrate ways in which the data's analytic potential can be enhanced. The joint Bureau of the Census and University of Wisconsin-Madison project to create public-use samples from the 1940 and 1950 Censuses of Population and Housing, and the plans of social scientists to publish a series of monographs on the use of the 1980 U.S. Census, are examples of how this broad strategy can be implemented.[53]

Some of the problems associated with the use of social science methodological and policy research can be traced to the fact that researchers are not part of the policy formation activities of government. There must be a greater level of communication between researchers and policy makers. Involvement in the internal decision-making process will indirectly improve the utilization of their research and policy recommendations. Along with administrators and policy makers, social researchers can assess research needs and examine the relationship of the statistical system to research activities outside the government. They can apply their training in organizational theory and public administration toward improving information management activities in government and toward strengthen-

ing the coordinating and overseeing functions that different parts of the federal statistical system perform:

> Closer ties between data producers and analysts will result in data that are more relevant to policy issues...[and] will also improve the quality of both data and analysis. Producers of data will have more direct feedback on quality from major users of data....Users will come to have a better understanding of the operational problems of collecting and processing data, and will design and perform their analyses with a better understanding of the limitations of the data.[54]

Maximize Capital Investment in Data Resources

Existing information structures are a vehicle for providing access to statistical data, for carrying out secondary analysis, or for linking the data to new collection activities. By providing public access and planning for future policy, teaching or research efforts can become more efficient. Further, data repositories, which generate cumulative knowledge activities, serve as laboratories for investigating methodological, analytical and policy problems. Data repositories are a more cost-effective and efficient way of reducing the structural problems inherent in government agency dissemination and preservation of computer-readable data.[55] The data library or archive is a way of maximizing the initial costly investment in producing data resources because it uses existing library and archival structures and information networks. Expensive start-up costs for creating new services can be avoided. Significant efficiencies take place for the user community by centralizing information about access to data by pooling resources and by using staff technical expertise.

The library environment also stimulates improvements in the quality of data. Common access creates a "commonality of research...among widely separated scholars."[56] The library acts as a scientific laboratory which encourages the sharing of data, multidisciplinary exploitation of evidence, and "multiple and...complex analytic applications...."[57] The data center makes a pedagogical contribution by allowing the student to participate in scientific inquiry and to develop problem-solving techniques and behavior. A recently completed study of factors influencing the sharing of computer-based resources for higher education shows a direct connection between utilization and sharing. It suggests that the "seemingly indirect attempts to broaden 'computer literacy' and computer use might have systematic effects on the level and nature of computer-based sharing."[58]

Less obviously, the data facility plays a role as an agent for assessing information transfer activities. It offers administrators and researchers the

opportunity to assess the technical, administrative, economic, and policy issues related to standards of data quality, documentation, access, and distribution.

At the same time, data must be more easily accessible to individuals and institutions without the resources to support large-scale data services for teaching and research. Public policy must foster development of small-scale data services within an institutional setting where it is not economically feasible to develop large-scale computational facilities or to rely upon highly trained support specialists in data and computation.[59]

Implied here is use of coordinating structures for *resource sharing*. The pressures of data growth and fiscal stress are real. Current economic and political realities have constrained efficient use of data and of the modern computer and communications technologies. Solutions must be sought on an organized national level for resource- and cost sharing, although individual institutions and organizations will make their own decisions appropriate for their local context.[60] Resource sharing means that the legal-institutional framework must be modified to reflect efficiencies that technological advances make possible. New methodologies for servicing the data user must be developed to take advantage of the benefits of technology. A different "mental set" is required, one that deemphasizes parochial concerns and emphasizes system-wide concerns; the information consumer will be the beneficiary, but so will the information center providing the service. It will however be particularly difficult to acquire this new mental set in a period of fiscal stress when the temptation is not to seek innovative, creative and flexible solutions[61] which appear to be acceptable only when there are slack resources. Nevertheless, the effort must be made because resources are becoming increasingly scarcer.

Educate the Information Specialist and End-User

Implied in capitalizing on the existing information transfer process is the need for strategies to improve the quality of education in professional library schools. Providing statistical data services requires a multidisciplinary approach so that, in addition to the traditional curriculum emphasizing organization and management tasks, analytic (intellectual) and technical enrichment is provided. A new process-oriented curriculum model (the communications model) must become the guiding principle underlying the intellectual and technical skills which the library or information science professional can apply.[62] More emphasis must be placed on use (knowledge creation) and the complex set of interdependencies created when people, information and technology interact. As Martyn

notes, there needs to be a deeper understanding of the ways in which information is acquired and used.[63] The information system must be redesigned to match the needs of users; implicit in this is the necessity for better understanding the role that information plays in innovation in ongoing work.

It is evident that modifications in undergraduate education are needed.[64] New information technologies are modifying society, organizational and human relationships, the work environment, language, and culture.[65] Statistical information is increasingly the basis on which public policy decisions are made. To understand the nature and direction of technological change and to evaluate public policy decisions will require a far more literate and numerite society than there is now.

Modify Political Choices

Increasing use of statistical data will require modifying traditional public and private sector relationships and the underlying philosophical premise that government should be producer, transfer agent, provider, and analyst of all statistical data it has heretofore been responsible for. A more realistic view is that government no longer has adequate resources to continue providing all the information services that it has provided to date. A better, more realistic mix that recognizes an increasing role by the private sector will be necessary.[66] Some creative solutions are needed so that these two sectors work more closely to develop public policy, whereby the development of standards is flexible and appropriate to different uses of the data. Also, so that government continues to provide "seed money" to support research and development for tools and services that respond to the public interest and to develop new methodologies, data collection procedures and structures. It is doubtful that the private sector has the funding base or interest to support the development of innovative products and services when immediate application, use and returns on the capital investment are not assured. Altering the relationship will not be easy or accomplished without creating structural discontinuities.

Another political choice reflects the necessity of modifying the sets of political relationships between institutions and nations which currently limit the efficient and inexpensive transfer of data. Recent efforts by the U.S. Congress to decentralize the communications system and by West European countries to create integrated communications networks must be encouraged. Other efforts, along the lines of research organizations in West European countries, must be designed to encourage the sharing of data. Institutions must be modified to take advantage of centers of expertise

both within the country and outside it; the distribution of data by centralized, hierarchical data organizations no longer appears to be the most economical way to transfer all types of data and expertise. The technology offers a means to modify political relationships, but it is up to consumers to recognize the benefits of the technology and to communicate these benefits to their political representatives and institutional administrators.

CONCLUSION

This paper reflects the philosophy that increasing the use of statistical data first requires knowledge of the properties of information systems and the process of information transfer and knowledge creation, in which organizations and individuals participate and are intimately linked. The second step is to select appropriate strategies so that objectives are adapted to external demands, so that existing patterns of interdependencies are modified by the components of the infrastructure and by building coalitions of support, and so that different organizational environments are created through law, education and knowledge utilization. Understanding the environment of the statistical data system and developing broad-based, system-oriented strategies can then be translated into strategies to develop innovative tools, products and services. In the coming decades public and private decisions will increasingly depend on access to statistical data. Our energies must be directed toward developing creative and flexible solutions that will foster utilization of this important information resource.

ACKNOWLEDGMENTS

I gratefully acknowledge the important role that Kathleen Heim played as "gatekeeper," and both the references and information on the influence of marketing principles on libraries provided by Darlene Weingand, Leigh Estabrook, and Jane Carter-Robbins. Of course, I alone bear responsibility for the comments contained in this paper.

REFERENCES

1. Pfeffer, Jeffrey, and Salancik, Gerald R. *The External Control of Organizations: A Resource Dependence Perspective.* New York: Harper & Row, 1978, p. 3.
2. Ibid.
3. For an excellent description of the information industry, *see* Schiller, Anita R., and Schiller, Herbert I. "Who Can Own What America Knows." *The Nation* 234(17 April 1982):461-63.

4. Robbin, Alice. *Final Report to the National Science Foundation Division of Science Resources Studies.* Madison, Wis.: University of Wisconsin Data and Program Library Service, 1981.

5. David, Martin. "Access to Data: The Frustration and Utopia of the Researcher." *Review of Public Data Use* 8(Dec. 1980):327-37.

6. Robbin, Alice. "Strategies for Improving Utilization of Computerized Statistical Data by the Social Scientific Community." *Social Science Information Studies* 1(Jan. 1981):89-109.

7. Ibid., pp. 91-92.

8. Boruch, Robert F., et al., eds. *Reanalyzing Program Evaluations.* San Francisco, Calif.: Jossey-Bass, 1981; and Hyman, Herbert H. *Secondary Analysis of Sample Surveys.* New York: Wiley and Sons, 1972.

9. Burghardt, John A., and Geraci, Vincent J. "State and Local Annual Population Estimation Methods Employed by the Bureau of the Census." *Review of Public Data Use* 8(Dec. 1980):339-54.

10. Ibid.

11. Hausman, Jerry A., and Trimble, John L. "Sample Design Considerations and Sample Selection for the Vermont Time-of-Day Electricity Use Survey." *Review of Public Data Use* 9(July 1981):83-86; and Hirst, Eric, and Blue, Jackalie L. "Federal-State Cooperation on Managing Disaggregate Energy Consumption Data." *Review of Public Data Use* 9(Nov. 1981):175-88.

12. Tweed, Dan L., et al. "Labor Force Deconcentration in the United States: An Examination of the Relative Impacts of Intrasystemic and Intersystemic Movement." *Review of Public Data Use* 9(July 1981):133-42.

13. Dale, Charles, et al. "Measuring Export Prices." *Review of Public Data Use* 9(Nov. 1981):199-209.

14. Ahlburg, Dennis, et al. "The Outlook for Higher Education: A Cohort Size Model of Enrollment of the College Age Population, 1948-2000." *Review of Public Data Use* 9(Nov. 1981):211-27.

15. Dodge, Richard W. and Lentzner, Harold R. "Patterns of Personal Series Incidents in the National Crime Survey." *Review of Public Data Use* 8(June 1980):27-41.

16. Behrens, John O. "Census Bureau Contributions to a Land Parcel Identification System." *Review of Public Data Use* 8(Aug. 1980):147-65.

17. Scheuren, Frederick, et al. "Exact Match Research Using the March 1973 Current Population Survey—Initial States" (HEW Publication no. (SSA)76-11750). Washington, D.C.: HEW, Social Security Administration, Office of Research and Statistics, 1976.

18. Robbin, Alice. "Technical Guidelines for Preparing and Documenting Data." In *Reanalyzing Program Evaluations,* edited by Robert F. Boruch, et al., pp. 84-143. San Francisco, Calif.: Jossey-Bass, 1981.

19. La Cour, A. "Policy and Recent Development in EUROSTAT—Methods of Disseminating Statistics." *Statistisk Tidskrift/Statistical Review* 15(1977):333-47. Reprinted in *European Political Data Newsletter* 35(June 1980):7-16.

20. Gonzales, Maria E., et al. "Federal Statistics from Administrative Records: Quality Issues" (paper prepared for the annual meeting of the American Association for the Advancement of Science, 4 Jan. 1982, Washington, D.C.).

21. Turner, Lloyd A. "The Value of Integrated Data Systems for Microanalysis of Social Service Programs." *Review of Public Data Use* 8(Aug. 1980):191-202.

22. Verway, David I. "A Critical Examination of the Dun and Bradstreet Data Files." *Review of Public Data Use* 8(Dec. 1980):369-74.

23. Ibid., p. 370.

24. David, Martin, and Robbin, Alice. "The Great Rift: Gaps Between Administrative Records and Knowledge Created Through Secondary Analysis." *Review of Public Data Use* 9(Nov. 1981):153-56; and David, Martin, et al. *Archival Preservation of Machine-Readable Records: The Final Report of the Wisconsin Survey of Machine-Readable Public Records.* Madison, Wis.: The State Historical Society of Wisconsin, 1981.

25. The President's Reorganization Project for the Federal Statistical System. "Improving the Federal Statistical System: Issue and Options." *Statistical Reporter* 81(Feb. 1981):133-

221; and Bailar, Barbara A., and Lanphier, Michael. *Development of Survey Methods to Assess Survey Practices*. Washington, D.C.: American Statistical Association, 1978.

26. Heim, Kathleen. "The Social Science Data Archives: A User Study." Ph.D. diss., University of Wisconsin School of Library Science, 1980.

27. Jones, Ray. "The Data Library in the University of Florida Libraries." *Library Trends* 30(Winter 1982):383-96; Ruus, Laine G.M. "The University of British Columbia Data Library: An Overview." *Library Trends* 30(Winter 1982):397-406; and Adams, Margaret O'Neill. "Online Numeric Data-Base Systems: A Resource for the Traditional Library." *Library Trends* 30(Winter 1982):435-54.

28. Robbin, Alice. *A Survey of Users of Scientific and Technical Manpower Data*. Madison, Wis.: University of Wisconsin, 1981.

29. Jones, "Data Library in University of Florida Libraries"; and Ruus, "University of British Columbia Data Library."

30. Robbin, Alice. "The Data and Program Library Service: A Case Study in Organizing Special Libraries for Computer-Readable Statistical Data." *Library Trends* 30(Winter 1982):407-34.

31. Rheinhold, Robert. "Data on Cuts Imperiled by Cuts in Data." *The New York Times*, 18 Oct. 1981, p. E4.

32. Crittenden, Ann. "A World with Fewer Numbers." *The New York Times*, 11 July 1982, p. F4; Rheinhold, "Data on Cuts Imperiled by Cuts in Data"; and Perlez, Jane. "Moynihan Cites New Loss in U.S. Funds." *The New York Times*, 26 July 1982, p. B2.

33. Perlez, "Moynihan Cites New Loss in U.S. Funds." *See also*, Duncan, Joseph W. "The Demand for Regional and Local-Area Statistics: Issues Concerning the National Response." *Statistical Reporter* 78(Jan. 1978):97-108.

34. Crittenden, "World with Fewer Numbers."

35. Ibid.

36. Ibid.

37. Ibid.; and Gonzales, et al. "Federal Statistics from Administrative Records."

38. Pollack, Andrew. "Digest's Disappointing Source." *The New York Times*, 26 July 1982, p. D1.

39. Clines, Francis X. "On the Prowl for Answers in a Jungle of Information." *The New York Times*, 14 Dec. 1980, p. 44. *See* for an informative description of the burgeoning repackaging of government information.

40. Duncan, Joseph W., et al. "Private Versus Public Sector Responsibility for the Collection, Distribution, and Analysis of Statistical Data." *Review of Public Data Use* 8(Dec. 1980):307-25.

41. Boruch, Robert F., and Cecil, Joe S. *Assuring the Confidentiality of Social Research Data*. Philadelphia: University of Pennsylvania Press, 1979.

42. The Privacy Protection Study Commission. *Personal Privacy in an Information Society*. Washington, D.C.: USGPO, 1977; Boruch, and Cecil, *Assuring Confidentiality of Social Research*; and Robbin, Alice. "Ethical Standards and Data Archives." *New Directions for Program Evaluation* 4(1978):7-18.

43. "Scholars in the Middle of Battles over Information." *The New York Times*, 11 July 1982, p. E23.

44. Rowe, Judith. "Online or On Tape." *IASSIST Newsletter* vol. 5, no. 4, 1982, pp. 79-84.

45. Robbin, Alice. "A Profile of Data Preservation Activities in University Libraries and Archives." *IASSIST Newsletter* vol. 5, no. 1, 1982, pp. 3-22.

46. President's Reorganization Project, "Improving the Federal Statistical System."

47. Rockwell, Richard C. "Notes on the Social Sciences as Producers of Technologies." *IASSIST Newsletter* vol. 5, no. 3, 1982, pp. 49-55.

48. Adams, "Online Numeric Data-Base Systems."; and Robbin, "Profile of Data Preservation Activities."

49. Brown, Maryann K. "Information for Planning." *Journal of Library Administration* 2(Summer/Fall/Winter 1981):187-215. Brown presents an excellent discussion on this subject.

50. *See* Bailar and Lanphier, *Development of Survey Methods.*

51. *See* Robbin, "Profile of Data Preservation Activities," fn. 45; Roistacher, Richard C., et al. *A Style Manual for Machine-Readable Data Files and Their Documentation.* Washington, D.C.: U.S. Department of Justice, Justice Statistics Bureau, 1980; and Dodd, Sue A. "Toward Integration of Catalog Records on Social Science Machine-Readable Data Into Existing Bibliographic Utilities: A Commentary." *Library Trends* 30(Winter 1982):335-61.

52. President's Reorganization Project, "Improving the Federal Statistical System," p. 172.

53. Winsborough, Halliman H. "Bureaucrats, Pedagogues and the Production of Data" (paper presented at the annual meetings of the American Association for the Advancement of Science, 4 Jan. 1982, Washington, D.C.); and Mare, Robert D. "Public Use Samples for the Past: A Progress Report on Constructing Public Use Samples From the 1940 and the 1950 U.S. Censuses of Population" (prepared for the Second International Conference of the International Association for Social Science Information Services, 9 May 1979, Ottawa, Canada).

54. President's Reorganization Project, "Improving the Federal Statistical System," p. 168.

55. Hofferbert, Richard I., and Clubb, Jerome M. "Introduction." *American Behavorial Scientist* 19(March/April 1976):381-86.

56. Miller, Warren E. "The Less Obvious Functions of Archiving Survey Research Data." *American Behavioral Scientist* 19(March/April 1976):411.

57. Hofferbert, and Clubb, "Introduction," p. 383.

58. Mebane, D.D., and Mebane, R.M. *Factors Influencing the Sharing of Computer-Based Resources for Higher Education and Research.* Princeton, N.J.: EDUCOM, 1981, pp. i-iv.

59. Heim, Kathleen M., ed. "Data Libraries for the Social Sciences." *Library Trends,* 30(Winter 1982):321-509.

60. Kronick, David A. "Goodbye to Farewells—Resource Sharing and Cost Sharing." *The Journal of Academic Librarianship* 8(July 1982):132-36.

61. Levine, Charles H., ed. *Fiscal Stress and Public Policy.* Beverly Hills, Calif.: Sage Publications, 1980.

62. Saracevic, Tefko. "A Model of Knowledge Communication as a Unifying Framework for Education in Librarianship and Information Science." *Information Reports and Bibliographies* vol. 7, no. 4, 1978, pp. 26-34.

63. Martyn, John. "Information Needs and Uses." In *Annual Review of Information Science and Technology,* vol. 9, edited by Carlos A. Curada, et al., pp. 3-23. Washington, D.C.: American Society for Information Science, 1974.

64. Kronick, "Goodbye to Farewells."

65. Nora, Simon, and Minc, Alain. *The Computerization of Society (L'informatisation de la societe).* Cambridge, Mass.: MIT Press, 1976.

66. Kasputys, Joseph E. "Private and Public Sector Responsibility for the Collection, Distribution and Analysis of Statistical Data." *IASSIST Newsletter* vol. 4, nos. 3/4, pp. 69-72.

VITA

Alice Robbin is head of the Data and Program Library Service at the University of Wisconsin-Madison. She has written extensively on the need to develop an information infrastructure and strategies for improving access to statistical data for social research. She is past president of IASSIST, the International Association for Social Sciences Information Service and Technology. Ms. Robbin has recently completed an investigation of the effect of privacy, confidentiality and open records access to

confidential records produced by state government. Ms. Robbin obtained her bachelor's degree in Political Science at Douglass College (Rutgers University) in 1966, her master's from the University of Wisconsin-Madison in 1969, and is currently completing work on her doctorate at the University of Wisconsin-Madison.